Search Results

Search Results for: **[distance<near>function* and decision<near>tree* and node* and gini<near>index*]**
Found **6** of **120,398 searched.**

Search within Results

GO > Advanced Search

> Search Help/Tips

Sort by: Title Publication Publication Date Score Binder

Results 1 - 6 of 6 short listing

1 Scalable algorithms for mining large databases 100%
Rajeev Rastogi , Kyuseok Shim
Tutorial notes of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining August 1999

2 Privacy-preserving data mining 89%
Rakesh Agrawal , Ramakrishnan Srikant
ACM SIGMOD Record , Proceedings of the 2000 ACM SIGMOD international conference on Management of data May 2000
Volume 29 Issue 2

A fruitful direction for future data mining research will be the development of techniques that incorporate privacy concerns. Specifically, we address the following question. Since the primary task in data mining is the development of models about aggregated data, can we develop accurate models without access to precise information in individual data records? We consider the concrete case of building a decision-tree classifier from training data in which the values of individual records have ...

3 Classification and regression: money *can* grow on trees 74%
Johannes Gehrke , Wie-Yin Loh , Raghu Ramakrishnan
Tutorial notes of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining August 1999
With over 800 million pages covering most areas of human endeavor, the World-wide Web is a fertile ground for data mining research to make a difference to the effectiveness of information search. Today, Web surfers access the Web through two dominant interfaces clicking on hyperlinks and searching via keyword queries This process is often tentative and unsatisfactory Better support is needed for expressing one's information need and dealing with a search result in more structured ways than ...

4 Web clustering: Evaluation of hierarchical clustering algorithms for document datasets 55%

Ying Zhao , George Karypis

Proceedings of the eleventh international conference on Information and knowledge management November 2002

Fast and high-quality document clustering algorithms play an important role in providing intuitive navigation and browsing mechanisms by organizing large amounts of information into a small number of meaningful clusters. In particular, hierarchical clustering solutions provide a view of the data at different levels of granularity, making them ideal for people to visualize and interactively explore large document collections.In this paper we evaluate different partitional and agglomerative approa ...

5 A new approach for evolving clusters 45%

Robert E. Marmelstein , Gary B. Lamont

Proceedings of the 1999 ACM symposium on Applied computing February 1999

6 Poster papers: Visualization support for a user-centered KDD process 0%

TuBao Ho , TrongDung Nguyen , DungDuc Nguyen

Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining July 2002

Viewing knowledge discovery as a user-centered process that requires an effective collaboration between the user and the discovery system, our work aims to support an active role of the user in that process by developing synergistic visualization tools integrated in our discovery system D2MS. These tools provide an ability of visualizing the entire process of knowledge discovery in order to help the user with data preprocessing, selecting mining algorithms and parameters, evaluating and comparin ...

Results 1 - 6 of 6 short listing

Your search matched **1** of **961219** documents.

A maximum of **1** results are displayed, **15** to a page, sorted by **Relevance** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

gini<near>index* <and> decision<near>tree*

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

1 **Scalable classifiers with dynamic pruning**
Gupta, S.K.; Somayajulu, D.V.L.N.; Arora, J.K.; Vasudha, B.;
Database and Expert Systems Applications, 1998. Proceedings. Ninth International Workshop on , 26-28 Aug. 1998
Page(s): 246 -251

[Abstract] [PDF Full-Text (212 KB)] **IEEE CNF**

JBK 8.15.3

Your search matched **1** of **961219** documents.

A maximum of **1** results are displayed, **15** to a page, sorted by **Relevance** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again.**

distance <near> function* <and> decision <near> tree* <and> node*

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

1 **Optimal partitioning for classification and regression trees**
Chou, P.A.;
Pattern Analysis and Machine Intelligence, IEEE Transactions on ,
Volume: 13 Issue: 4 , April 1991
Page(s): 340 -354

[Abstract] [PDF Full-Text (1308 KB)] **IEEE JNL**

Search Results

Search Results for: **[decision <near> tree <paragraph> partition* <paragraph> gini <near> index*]**
Found **7** of **120,398 searched.**

Search within Results

GO > Advanced Search

> Search Help/Tips

Sort by: Title Publication Publication Date Score Binder

Results 1 - 7 of 7 short listing

1 Research papers: data mining: An integrated approach for scaling up classification and prediction algorithms for data mining 100%

Patricia E. N. Lutu

Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology September 2002

Classification and prediction algorithms for machine learning typically require all training data to be resident in memory during decision tree construction. Typically, a flat file is created from database or data warehouse data and loaded into memory for processing. This severely limits the scalability of these algorithms to practical data mining tasks. Some attempts have been made by researchers to implement disk-based algorithms which can handle much larger training sets. Both approaches suff ...

2 Data exploration: HD-Eye: visual clustering of high dimensional data 100%

Alexander Hinneburg , Daniel A. Keim , Markus Wawryniuk

Proceedings of the 2002 ACM SIGMOD international conference on Management of data June 2002

Clustering of large data bases is an important research area with a large variety of applications in the data base context. Missing in most of the research efforts are means for guiding the clustering process and understanding the results, which is especially important for high dimensional data. Visualization technology may help to solve this problem since it provides effective support of different clustering paradigms and allows a visual inspection of the results. The *HD-Eye* (high-dim. e ...

3 Classification: SQL database primitives for decision tree classifiers 100%

Kai-Uwe Sattler , Oliver Dunemann

Proceedings of the tenth international conference on Information and knowledge management October 2001

Scalable data mining in large databases is one of today's challenges to database technologies.

Thus, substantial effort is dedicated to a tight coupling of database and data mining systems leading to database primitives supporting data mining tasks. In order to support a wide range of tasks and to be of general usage these primitives should be rather building blocks than implementations of specific algorithms. In this paper, we describe primitives for building and applying decision tree classifi ...

4 Data Mining with optimized two-dimensional association rules 100%

Takeshi Fukuda , Yasuhiko Morimoto , Shimichi Morishita , Takeshi Tokuyama

ACM Transactions on Database Systems (TODS) June 2001

Volume 26 Issue 2

We discuss data mining based on association rules for two numeric attributes and one Boolean attribute. For example, in a database of bank customers, Age and Balance are two numeric attributes, and CardLoan is a Boolean attribute. Taking the pair (Age, Balance) as a point in two-dimensional space, we consider an association rule of the form Age,Balance ∈P⇒

5 Towards an effective cooperation of the user and the computer for classification 100%

Mihael Ankerst , Martin Ester , Hans-Peter Kriegel

Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining August 2000

6 Privacy-preserving data mining 100%

Rakesh Agrawal , Ramakrishnan Srikant

ACM SIGMOD Record , Proceedings of the 2000 ACM SIGMOD international conference on Management of data May 2000

Volume 29 Issue 2

A fruitful direction for future data mining research will be the development of techniques that incorporate privacy concerns. Specifically, we address the following question. Since the primary task in data mining is the development of models about aggregated data, can we develop accurate models without access to precise information in individual data records? We consider the concrete case of building a decision-tree classifier from training data in which the values of individual records have ...

7 Towards on-line analytical mining in large databases 100%

Jiawei Han

ACM SIGMOD Record March 1998

Volume 27 Issue 1

Great efforts have been paid in the Intelligent Database Systems Research Lab for the research and development of efficient data mining methods and construction of on-line analytical data mining systems.Our work has been focused on the integration of data mining and OLAP technologies and the development of scalable, integrated, and multiple data mining functions. A data mining system, DBMiner, has been developed for interactive mining of multiple-level knowledge in large relational databases and ...

Search Results

Search Results for: **[record* <paragraph> class* <paragraph> gini <near> index]**
Found **4** of **120,398 searched.**

Search within Results

GO > Advanced Search

> Search Help/Tips

Sort by: Title Publication Publication Date Score Binder

Results 1 - 4 of 4 short listing

1 Classification: SQL database primitives for decision tree classifiers 100%
Kai-Uwe Sattler , Oliver Dunemann
Proceedings of the tenth international conference on Information and knowledge management October 2001
Scalable data mining in large databases is one of today's challenges to database technologies. Thus, substantial effort is dedicated to a tight coupling of database and data mining systems leading to database primitives supporting data mining tasks. In order to support a wide range of tasks and to be of general usage these primitives should be rather building blocks than implementations of specific algorithms. In this paper, we describe primitives for building and applying decision tree classifi ...

2 Towards an effective cooperation of the user and the computer for classification 100%
Mihael Ankerst , Martin Ester , Hans-Peter Kriegel
Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining August 2000

3 Privacy-preserving data mining 100%
Rakesh Agrawal , Ramakrishnan Srikant
ACM SIGMOD Record , Proceedings of the 2000 ACM SIGMOD international conference on Management of data May 2000
Volume 29 Issue 2

A fruitful direction for future data mining research will be the development of techniques that incorporate privacy concerns. Specifically, we address the following question. Since the primary task in data mining is the development of models about aggregated data, can we develop accurate models without access to precise information in individual data records? We consider the concrete case of building a decision-tree classifier from training data in which the values of individual records have ...

4 Learning decision tree classifiers 100%
J. R. Quinlan
ACM Computing Surveys (CSUR) March 1996
Volume 28 Issue 1

Results 1 - 4 of 4 short listing

Search Results

Search Results for: **[distance<near>function* and decision<near>tree* and node*]**
Found **972** of **120,398 searched.**

Warning: Maximum result set of 200 exceeded. Consider refining.

Search within Results

Sort by: Title Publication Publication Date Score

Results 1 - 20 of 200 short listing

Prev Page 1 2 3 4 5 6 7 8 9 10 Next Page

1 Clustering: ReCoM: reinforcement clustering of multi-type interrelated data objects 100%
Jidong Wang , Huajun Zeng , Zheng Chen , Hongjun Lu , Li Tao , Wei-Ying Ma
Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval July 2003
Most existing clustering algorithms cluster highly related data objects such as Web pages and Web users separately. The interrelation among different types of data objects is either not considered, or represented by a static feature space and treated in the same ways as other attributes of the objects. In this paper, we propose a novel clustering approach for clustering multi-type interrelated data objects, ReCoM (Reinforcement Clustering of Multi-type Interrelated data objects). Under this appr ...

2 Fast detection of communication patterns in distributed executions 100%
Thomas Kunz , Michiel F. H. Seuren
Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research November 1997
Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

3 A retrospective on constraint databases 100%
Peter Revesz

Proceedings of the Paris C. Kanellakis memorial workshop on Principles of computing & knowledge: Paris C. Kanellakis memorial workshop on the occasion of his 50th birthday June 2003

In this paper we give a review of constraint databases, a field that was started by Paris Kanellakis, Gabriel Kuper and the author. The review includes basic concepts of data representation, constraint query languages, and query evaluation. We also illustrate applications of constraint databases in the areas of model checking, data mining, trust management, Diophantine polynomial equations, and moving objects.

4 Data structures: Proximate planar point location 100%

John Iacono , Stefan Langerman

Proceedings of the nineteenth conference on Computational geometry June 2003

A new data structure is presented for planar point location that executes a point location query quickly if it is spatially near the previous query. Given a triangulation T of size n and a sequence of point location queries $A=q_1, q_m$, the structure presented executes q_i in time $O(\log d(q_{i-1},q_i))$. The distance function, d, that is used is a two dimensional generalization of rank distance that counts th ...

5 A survey on wavelet applications in data mining 100%

Tao Li , Qi Li , Shenghuo Zhu , Mitsunori Ogihara

ACM SIGKDD Explorations Newsletter December 2002

Volume 4 Issue 2

Recently there has been significant development in the use of wavelet methods in various data mining processes. However, there has been written no comprehensive survey available on the topic. The goal of this is paper to fill the void. First, the paper presents a high-level data-mining framework that reduces the overall process into smaller components. Then applications of wavelets for each component are reviewd. The paper concludes by discussing the impact of wavelets on data mining research an ...

6 Contributed articles on online, interactive, and anytime data mining: Towards effective and interpretable data mining by visual interaction 100%

Charu C. Aggarwal

ACM SIGKDD Explorations Newsletter January 2002

Volume 3 Issue 2

The primary aim of most data mining algorithms is to facilitate the discovery of concise and interpretable information from large amounts of data. However, many of the current formalizations of data mining algorithms have not quite reached this goal. One of the reasons for this is that the focus on using purely automated techniques has imposed several constraints on data mining algorithms. For example, any data mining problem such as clustering or association rules requires the specification of ...

7 Contributed articles on online, interactive, and anytime data mining: Mining data streams under block evolution 100%

Venkatesh Ganti , Johannes Gehrke , Raghu Ramakrishnan

ACM SIGKDD Explorations Newsletter January 2002

Volume 3 Issue 2

In this paper we survey recent work on incremental data mining model maintenance and change detection under *block evolution.* In block evolution, a dataset is updated periodically

through insertions and deletions of *blocks* of records at a time. We describe two techniques: (1) We describe a generic algorithm for model maintenance that takes any traditional incremental data mining model maintenance algorithm and transforms it into an algorithm that allows restrictions on a temporal su ...

8 Estimating business targets 100%

Piew Datta , James H. Drew , Andrew Betz , D. R. Mani , Jeffery Howard

Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining August 2001

Determining and setting maximal revenue expectations or other business performance targets---whether it is for regional company divisions or individual customers---can have profound financial implications. Operational techniques are changed, staffing levels are altered and management attention is re-focused---all in the name of expectations. In practice these expectations are often derived in an ad hoc manner. To address this unsupervised task, we combine nearest neighbor methods and classical s ...

9 Computational geometry 100%

Joseph S. B. Mitchell , Joseph O'Rourke

ACM SIGACT News September 2001

Volume 32 Issue 3

A compendium of thirty previously published open problems in computational geometry is presented

10 Sampling algorithms: lower bounds and applications 100%

Ziv Bar-Yossef , Ravi Kumar , D. Sivakumar

Proceedings of the thirty-third annual ACM symposium on Theory of computing July 2001

We develop a framework to study probabilistic sampling algorithms that approximate general functions of the form \genfunc, where \domain and \range are arbitrary sets. Our goal is to obtain lower bounds on the query complexity of functions, namely the number of input variables x_i that any sampling algorithm needs to query to approximate f(x_1,\ldots,x_n).We define two quantitative properties of ...

11 An Algorithm for Finding Best Matches in Logarithmic Expected Time 100%

Jerome H. Freidman , Jon Louis Bentley , Raphael Ari Finkel

ACM Transactions on Mathematical Software (TOMS) September 1977

Volume 3 Issue 3

12 Identifying prospective customers 100%

Paul B. Chou , Edna Grossman , Dimitrios Gunopulos , Pasumarti Kamesam

Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining August 2000

13 Efficient algorithms for mining outliers from large data sets 100%

Sridhar Ramaswamy , Rajeev Rastogi , Kyuseok Shim

ACM SIGMOD Record , Proceedings of the 2000 ACM SIGMOD international conference on Management of data May 2000

Volume 29 Issue 2

In this paper, we propose a novel formulation for distance-based *outliers* that is based on the distance of a point from its k^{th} nearest neighbor. We rank each point on the basis of its distance to its k^{th} nearest neighbor and declare the top n points in this ranking to be outliers. In addition to developing relatively straightforward solutions to finding such outliers based on the classical nested-loop join and index join algorithms, we develo ...

14 XTRACT: a system for extracting document type descriptors from XML documents 100%
Minos Garofalakis , Aristides Gionis , Rajeev Rastogi , S. Seshadri , Kyuseok Shim
ACM SIGMOD Record , Proceedings of the 2000 ACM SIGMOD international conference on Management of data May 2000
Volume 29 Issue 2

XML is rapidly emerging as the new standard for data representation and exchange on the Web. An XML document can be accompanied by a *Document Type Descriptor* (DTD) which plays the role of a schema for an XML data collection. DTDs contain valuable information on the structure of documents and thus have a crucial role in the efficient storage of XML data, as well as the effective formulation and optimization of XML queries. In this paper, we propose XTRACT, a novel system for inferring a ...

15 Data clustering: a review 100%
A. K. Jain , M. N. Murty , P. J. Flynn
ACM Computing Surveys (CSUR) September 1999
Volume 31 Issue 3
Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic co ...

16 Lower Bounds for Selection in X + Y and Other Multisets 100%
Donald B. Johnson , Samuel D. Kashdan
Journal of the ACM (JACM) October 1978
Volume 25 Issue 4

17 Parametric Combinatorial Computing and a Problem of Program Module Distribution 100%
Dan Gusfield
Journal of the ACM (JACM) July 1983
Volume 30 Issue 3

18 An optimal algorithm for approximate nearest neighbor searching 100%
Sunil Arya , David M. Mount , Nathan S. Netanyahu , Ruth Silverman , Angela Wu
Proceedings of the fifth annual ACM-SIAM symposium on Discrete algorithms January 1994

19 Fuzzy distances and image processing 100%
Isabelle Bloch , Henri Maitre
Proceedings of the 1995 ACM symposium on Applied computing February 1995

20 Lower bounds for high dimensional nearest neighbor search and related problems 100%
Allan Borodin , Rafail Ostrovsky , Yuval Rabani
Proceedings of the thirty-first annual ACM symposium on Theory of computing May 1999

Results 1 - 20 of 200 short listing

Prev Page 1 2 3 4 5 6 7 8 9 10 Next Page
